離散時間マルコフ系における最大原理について (動的計画法の研究会報告集)

AUTHOR(S):

古川, 長太

CITATION:

古川, 長太. 離散時間マルコフ系における最大原理について (動的計画法の研究会報告集). 数理解析研究所講究録 1969, 67: 13-24

ISSUE DATE:

1969-02

URL:

http://hdl.handle.net/2433/107897

数理解析研究所講究録
67巻 1969年 13-24

離散時間マルコフ系における最大原理について

九州大学 理学部 古川長太

§1. 序

制御項を含んだ常微分方程式系や差分方程式系において，評価汎関数を最大または最小ならしめる最適制御の問題に対しては，最適制御のための必要条件としての最大原理や，最適制御の存在のための十分条件などが，多くの人達によって研究され，事実，それらについての具体的な結果が，個々の問題について詳細に，導かれるに至っている。

一方，Howard, Blackwell で代表されるマルコフ型決定過程における最適化問題に対しては，一般に，ある意味での最適解の存在と最適解の性質が知られているが，最適解の構成に関しては，state, action 共に高々可付番個の場合に限って Howard, Derman などの研究が有るのみである。

マルコフ型決定過程で特に有限 stage に限定し，従って，評価汎関数も有限 stage にわたる利益の期待値の総和と云う

杯なものにすると，これは正に，マルコフ過程と云うNoiseを含んだ確率的差分方程式系における time fixed の最適制御の問題になる。これに関しては，従来，Bellman の Dynamic Programming formulation による解析が行われて来たが（理論的に explicit な型で解けるとは限らないが），始めに述べた最適制御の問題との関連において，この system での最適制御のための最大原理について考察してみることは，最適解の構成のための何等かのヒントとしての意味を持つものである。この故に，この報告では，有限 stage の Blackwell model を含むある一般的な確率的差分方程式系に対して，最大原理（正確には最大原理の類似）を導くことを目的とする。

§2. 諸定義と問題の定式化

ここでは，次の杯な確率的差分方程式系を扱う。

$$(2.1)\qquad \begin{cases} x_{j+1} = f(\xi_{j-1}, x_j, v_j, \xi_j), & j=1,2,\cdots,N-1, \\ x_1 = c \end{cases}$$

ただし

$x_j \in R^n$: time $t=j$ における n 次元 state vector

$v_j \in V_j \subset R^m$: time $t=j$ における m 次元 control action vector

V_j : time $t=j$ における feasible control action の集合

$\xi_j \in S \subset R^r$: time $t=j$ における r 次元 disturbance vector

ここに次の仮定をおく。

仮定 (I)

$\{\xi_j,\ 0 \leqq j \leqq N-1\}$ は state space S 上の離散時間マルコフ過程で，かつ $\{\xi_j\}$ の transition probability は定常でなく，一般に ξ_j から ξ_{j+1} への transition probability は time $t=j$ においてとられた control action (そしてそれのみに)に依存する。

x_N の第1成分を $x_N(0)$ で表わすことにし，制御の目的は $E\{x_N(0)\}$ を ξ_0 につき一様に最小ならしめる $\{u_1, u_2, \cdots, u_{N-1}\}$ (ただし u_j は time $t=j$ における control action を表わす関数)を求めることであるとする。ただし，ここでは次の諸定義に見られるように，$\{u_1, u_2, \cdots u_{N-1}\}$ の class を幾分か限定することにする。

定義 1　u_j を，$u_j(\xi_{j-1}, x_j) = v_j$ により定義される $S \times R^n \to V_j$ なる measurable mapping として，$u = \{u_1, u_2, \cdots u_{N-1}\}$ を admissible control と云う。

定義 2　admissible control の全体を Γ とする。control u^* が有って，u^* に対応する trajectory，(2.1) の解が $x^* = \{x_1^*, x_2^*, \cdots, x_N^*\}$ であるとすると，

$$E\{x_N^*(0)\} = \inf_{u \in \Gamma} E\{x_N(0)\} \qquad (\xi_0 \text{ に関し一様に})$$

なるとき，u^* を optimal control と云う。

仮定(II)　f は各 argument につき2回連続微分可能

仮定(III)　admissible control u に対応するマルコフ過程を $\{\xi_j\}$, admissible control $\hat{u} = u + \varepsilon\zeta$ (ただし ζ は有界) に対応するマルコフ過程を $\{\hat{\xi}_j\}$ とすると,

$$E\{|\xi_j - \hat{\xi}_j|\} = o(\varepsilon) \qquad (j=1,2,\cdots,N-1)。$$

§3. 最大原理

補題　仮定(I)の下で system (2.1) に対する optimal control を $u^* = \{u_1^*, u_2^*, \cdots, u_{N-1}^*\}$ とし, u^* に対応する (2.1) の解を $x^* = \{x_1^*, x_2^*, \cdots, x_N^*\}$, u^* に対応するマルコフ過程を $\xi^* = \{\xi_1^*, \xi_2^*, \cdots, \xi_{N-1}^*\}$ とすれば

(3.1)　$$E\{x_N(0) - x_N^*(0) \mid (\xi_{j-1}, x_j) = (\xi_{j-1}^*, x_j^*)\} \geqq 0 \quad a.s.$$

for $j=1,2,\cdots,N-1$,
for all ξ_0,

ただし, x_N は任意の admissible control に対応する (2.1) の解の time $t=N$ におけるもの,

が成立する。

(証明)

(3.1) が成立しないとする。即ち, i ($1 \leqq i \leqq N-1$), ξ_0, A ($A \subset S \times R^m$), admissible control S が存在して

(i)　$P_{\xi_0}\{(\xi_{i-1}^*, x_i^*) \in A\} > 0$

(ii)　control S に対応する trajectory を $\tilde{x}$, S に対応する

マルコフ過程を $\tilde{\xi} = \{\tilde{\xi}_1, \tilde{\xi}_2, \cdots, \tilde{\xi}_{N-1}\}$ とすれば

$$E_{\xi_0}\{\tilde{x}_N(0) - x^*_N(0) \mid (\tilde{\xi}_{i-1}, \tilde{x}_i) = (\xi^*_{i-1}, x^*_i) \in A\} < 0$$

が成立つ。

次に, admissible control t を定義する。 $t = \{t_1, t_2, \cdots, t_{N-1}\}$

$$t_j = \begin{cases} u^*_j & \text{if } j < i \\ u^*_j & \text{if } j \geq i, \ (\xi^*_{i-1}, x^*_i) \notin A \\ s_j & \text{if } j \geq i, \ (\xi^*_{i-1}, x^*_i) \in A . \end{cases}$$

故に $(\xi^*_{i-1}, x^*_i) \in A$ の条件の下では, 明らかに, $t = \{u^*_1, u^*_2, \cdots, u^*_{i-1}, s_i, s_{i+1}, \cdots, s_{N-1}\}$ となる。 t に対応する trajectory を $\hat{x} = \{\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_N\}$, t に対応するマルコフ過程を $\hat{\xi} = \{\hat{\xi}_1, \hat{\xi}_2, \cdots, \hat{\xi}_{N-1}\}$ とすると, x^*, $\hat{x}$ の定義より次の式が成立つ。

(3.2) $$x^*_{j+1} = f(\xi^*_{j-1}, x^*_j, u^*_j(\xi^*_{j-1}, x^*_j), \xi^*_j), \qquad j = 1, 2, \cdots, N-1,$$

(3.3) $$\hat{x}_{j+1} = f(\hat{\xi}_{j-1}, \hat{x}_j, t_j(\hat{\xi}_{j-1}, \hat{x}_j), \hat{\xi}_j), \qquad j = 1, 2, \cdots, N-1.$$

t の定義より, (3.3) の中, 特に次を得る。

$$\begin{aligned} \hat{x}_2 &= f(\hat{\xi}_0, \hat{x}_1, u^*_1(\hat{\xi}_0, \hat{x}_1), \hat{\xi}_1) \\ \hat{x}_3 &= f(\hat{\xi}_1, \hat{x}_2, u^*_2(\hat{\xi}_1, \hat{x}_2), \hat{\xi}_2) \\ &\vdots \qquad\qquad\qquad \vdots \\ \hat{x}_i &= f(\hat{\xi}_{i-2}, \hat{x}_{i-1}, u^*_{i-1}(\hat{\xi}_{i-2}, \hat{x}_{i-1}), \hat{\xi}_{i-1}) \end{aligned}$$

故に, (3.2), (3.3) は $j = 1, 2, \cdots, i-1$ に関しては equivalent equation である。 故に

$$\hat{\xi}_{i-1} = \xi^*_{i-1}, \qquad \hat{x}_i = x^*_{i-1}$$

$\hat{x}$ の定義より，(3.3) を用いて次を得る。

(3.4) $(\xi^*_{i-1}, x^*_i) \in A$ の条件の下で

$$\hat{x}_{j+1} = f(\hat{\xi}_{j-1}, \hat{x}_j, s_j(\hat{\xi}_{j-1}, \hat{x}_j), \xi_j) \qquad (j = i, i+1, \cdots, N-1)$$

一方，$\tilde{x}$ の定義より

(3.5) $$\tilde{x}_{j+1} = f(\tilde{\xi}_{j-1}, \tilde{x}_j, s_j(\tilde{\xi}_{j-1}, \tilde{x}_j), \tilde{\xi}_j), \qquad (j = i, i+1, \cdots, N-1)$$

故に (3.4), (3.5) より

(3.6) $$E\{\hat{x}_N(0) \mid (\xi^*_{i-1}, x^*_i) \in A\} = E\{\tilde{x}_N(0) \mid (\tilde{\xi}_{i-1}, \tilde{x}_i) = (\xi^*_{i-1}, x^*_i) \in A\}$$

次に

(3.7) $$E\{\hat{x}_N(0) \mid (\xi^*_{i-1}, x^*_i) \in A\} = E\{\hat{x}_N(0) \mid (\xi^*_{i-1}, x^*_i) = (\tilde{\xi}_{i-1}, \tilde{x}_i) \in A\}$$

故に (3.6), (3.7) より

(3.8) $$E\{\hat{x}_N(0) - \tilde{x}_N(0) \mid (\xi^*_{i-1}, x^*_i) = (\tilde{\xi}_{i-1}, \tilde{x}_i) \in A\} = 0$$

また，

$$\begin{aligned}
&E_{\xi_0}\{\hat{x}_N(0) - x^*_N(0) \mid (\xi^*_{i-1}, x^*_i) \in A\} \\
&= E_{\xi_0}\{\hat{x}_N(0) - x^*_N(0) \mid (\xi^*_{i-1}, x^*_i) = (\tilde{\xi}_{i-1}, \tilde{x}_i) \in A\} \\
&= E_{\xi_0}\{\hat{x}_N(0) - \tilde{x}_N(0) \mid (\xi^*_{i-1}, x^*_i) = (\tilde{\xi}_{i-1}, \tilde{x}_i) \in A\} \\
&\quad + E_{\xi_0}\{\tilde{x}_N(0) - x^*_N(0) \mid (\xi^*_{i-1}, x^*_i) = (\tilde{\xi}_{i-1}, \tilde{x}_i) \in A\} \\
&= E_{\xi_0}\{\tilde{x}_N(0) - x^*_N(0) \mid (\xi^*_{i-1}, x^*_i) = (\tilde{\xi}_{i-1}, \tilde{x}_i) \in A\} \quad (\because (3.8)) \\
&< 0 \qquad (\because (ii))
\end{aligned}$$

即ち，

(3.9) $$E_{\xi_0}\{\hat{x}_N(0) - x^*_N(0) \mid (\xi^*_{i-1}, x^*_i) \in A\} < 0$$

$$E_{\xi_0}\{\hat{x}_N(0)-x_N^*(0)\}=E_{\xi_0}\{\hat{x}_N(0)-x_N^*(0)\,|\,(\xi_{i-1}^*,x_i^*)\in A\}P_{\xi_0}\{(\xi_{i-1}^*,x_i^*)\in A\}$$
$$+E_{\xi_0}\{\hat{x}_N(0)-x_N^*(0)\,|\,(\xi_{i-1}^*,x_i^*)\notin A\}P_{\xi_0}\{(\xi_{i-1}^*,x_i^*)\notin A\}$$

上式右辺の第2項における $E_{\xi_0}\{\quad\}$ は A の定義により 0 であるから，(i) 及び (3.9) により

$$E_{\xi_0}\{\hat{x}_N(0)-x_N^*(0)\}<0$$
$$\therefore E_{\xi_0}\{\hat{x}_N(0)\}<E_{\xi_0}\{x_N^*(0)\}$$

これは，u^* が optimal なることに反する。故に (3.1) は成立する。

(補題証明終り)

u^* を optimal control とし，x^* を u^* に対応する trajectory とする。

F_j を次の式で定義する。

$$F_j \underset{n\times n}{\equiv} \frac{\partial f(\xi_{j-1},x_j,v_j,\xi_j)}{\partial x_j}+\frac{\partial f(\xi_{j-1},x_j,v_j,\xi_j)}{\partial v_j}\cdot\frac{\partial u_j}{\partial x_j}\Bigg|_{\substack{x_j=x_j^*,\ \xi_j=\xi_j^*,\\ u_j=u_j^*,\ \xi_{j-1}=\xi_{j-1}^*\\ v_j=v_j^*,}}$$

ただし，$v_j^*=u_j^*(\xi_{j-1}^*,x_j^*)$，$\xi^*$ は u^* に対応するマルコフ過程とする。

次に，stochastic n-vector p_j を定義する。

$$(3.10)\qquad \begin{cases} p_{j-1}^T = p_j^T F_j, & 2\leqq j\leqq N-1,\\ p_{N-1}^T=(1,0,0,\cdots,0)\end{cases}$$

更に，

$$H_j(p_j,\xi_{j-1},x_j,u_j,\xi_j)\equiv p_j^T f(\xi_{j-1},x_j,u_j,\xi_j)$$

とおく。

定理　仮定(I),(II),(III)を仮定する。

(2.1) に対する optimal control u^* が存在し，かつ，u^* は x に関し $(a.s.)$ 2回連続連微分可能であるとする。

$\Longrightarrow$

次の (i),(ii) を充たす (3.10) の解 $p_1, p_2, \cdots, p_{N-1}$ が存在する：

(i) $E\{H_{N-1}(p_{N-1}, \xi^*_{N-2}, x^*_{N-1}, u_{N-1}, \xi_{N-1}) \mid (\xi^*_{N-2}, x^*_{N-1})\}$

$\geq E\{H_{N-1}(p_{N-1}, \xi^*_{N-2}, x^*_{N-1}, u^*_{N-1}, \xi^*_{N-1}) \mid (\xi^*_{N-2}, x^*_{N-1})\}$ a.s.

(ii) $1 \leq j \leq N-2$ に対しては

$u^*_j + \varepsilon\zeta_j \in \Gamma_j$ (ただし ζ_j は有界)，$\varepsilon > 0$

$\Longrightarrow E\{H_j(p_j, \xi^*_{j-1}, x^*_j, u^*_j + \varepsilon\zeta_j, \xi_j) \mid (\xi^*_{j-1}, x^*_j)\} + o(\varepsilon)$

$\geq E\{H_j(p_j, \xi^*_{j-1}, x^*_j, u^*_j, \xi^*_j) \mid (\xi^*_{j-1}, x^*_j)\}$ a.s.

(証明)

$\varepsilon > 0$ とし，$\zeta = \{\zeta_1, \zeta_2, \cdots, \zeta_{N-1}\}$ を $u^*_j + \varepsilon\zeta_j \in \Gamma_j$ $(1 \leq j \leq N-1)$ なる有界な関数とする。

$r \equiv$ the largest integer i such that $1 \leq i \leq N-1$ and $P_r\{\zeta_i(\xi^*_{i-1}, x^*_i) = 0\} < 1$,

$\tilde{x} = \{\tilde{x}_0, \tilde{x}_1, \cdots, \tilde{x}_N\}$ を $u^* + \varepsilon\zeta$ に対応する trajectory とする。

Case $r = N-1$

$E\{\tilde{x}_N(\omega) - x^*_N(\omega) \mid (\tilde{\xi}_{N-2}, \tilde{x}_{N-1}) = (\xi^*_{N-2}, x^*_{N-1})\}$

$$= E\{ f^0(\tilde{\xi}_{N-2}, \tilde{x}_{N-1}, u^*_{N-1}(\tilde{\xi}_{N-2}, \tilde{x}_{N-1}) + \varepsilon\zeta_{N-1}(\tilde{\xi}_{N-2}, \tilde{x}_{N-1}), \tilde{\xi}_{N-1})$$
$$- f^0(\xi^*_{N-2}, x^*_{N-1}, u^*_{N-1}(\xi^*_{N-2}, x^*_{N-1}), \xi^*_{N-1}) \mid (\tilde{\xi}_{N-2}, \tilde{x}_{N-1}) = (\xi^*_{N-2}, x^*_{N-1}) \}$$

(ただし f^0 は f の第1成分)

$$= E\{ f^0(\xi^*_{N-2}, x^*_{N-1}, u^*_{N-1}(\xi^*_{N-2}, x^*_{N-1}) + \varepsilon\zeta_{N-1}(\xi^*_{N-2}, x^*_{N-1}), \tilde{\xi}_{N-1})$$
$$- f^0(\xi^*_{N-2}, x^*_{N-1}, u^*_{N-1}(\xi^*_{N-2}, x^*_{N-1}), \xi^*_{N-1}) \mid (\xi^*_{N-2}, x^*_{N-1}) \}$$

ここに，仮定より $P_r\{\zeta_{N-1}(\xi^*_{N-2}, x^*_{N-1}) = 0\} < 1$ だから，上式は恒等的に0ではない。更に，$f^0 = p^T_{N-1} f$ を用いて

$$= E\{ p^T_{N-1}[f(\xi^*_{N-2}, x^*_{N-1}, u^*_{N-1} + \varepsilon\zeta_{N-1}, \tilde{\xi}_{N-1})$$
$$- f(\xi^*_{N-2}, x^*_{N-1}, u^*_{N-1}, \xi^*_{N-1})] \mid (\xi^*_{N-2}, x^*_{N-1}) \}$$

故に補題により

$$E\{ p^T_{N-1} f(\xi^*_{N-2}, x^*_{N-1}, u^*_{N-1} + \varepsilon\zeta_{N-1}, \tilde{\xi}_{N-1}) \mid (\xi^*_{N-2}, x^*_{N-1}) \}$$
$$\geq E\{ p^T_{N-1} f(\xi^*_{N-2}, x^*_{N-1}, u^*_{N-1}, \xi^*_{N-1}) \mid (\xi^*_{N-2}, x^*_{N-1}) \} \quad a.s.$$

即ち

$$E\{ H_{N-1}(p_{N-1}, \xi^*_{N-2}, x^*_{N-1}, u^*_{N-1} + \varepsilon\zeta_{N-1}, \tilde{\xi}_{N-1}) \mid (\xi^*_{N-2}, x^*_{N-1}) \}$$
$$\geq E\{ H(p_{N-1}, \xi^*_{N-2}, x^*_{N-1}, u^*_{N-1}, \xi^*_{N-1}) \mid (\xi^*_{N-2}, x^*_{N-1}) \} \quad a.s.$$

これは，$u^*_{N-1} + \varepsilon\zeta_{N-1} \in \Gamma_{N-1}$ なるすべての $u^*_{N-1} + \varepsilon\zeta_{N-1}$ に対して成立。故に

$$E\{ H_{N-1}(p_{N-1}, \xi^*_{N-2}, x^*_{N-1}, u_{N-1}, \xi_{N-1}) \mid (\xi^*_{N-2}, x^*_{N-1}) \}$$
$$\geq E\{ H_{N-1}(p_{N-1}, \xi^*_{N-2}, x^*_{N-1}, u^*_{N-1}, \xi^*_{N-1}) \mid (\xi^*_{N-2}, x^*_{N-1}) \} \quad a.s.$$

Case $\tau = N-2$

admissible control $\hat{u} = \{\hat{u}_1, \hat{u}_2, \cdots, \hat{u}_{N-1}\}$ を次のように定義する。

$$\begin{cases} \hat{u}_j = u_j^* + \varepsilon \zeta_j, & 1 \leqq j \leqq N-2 \\ \hat{u}_{N-1} = u_{N-1}^*. \end{cases}$$

$\hat{u}$ に対応する trajectory を $\hat{x}$, マルコフ過程を $\hat{\xi}$ とする。

$$\begin{aligned} \hat{x}_N - x_N^* &= f(\hat{\xi}_{N-2}, \hat{x}_{N-1}, \hat{u}_{N-1}(\hat{\xi}_{N-2}, \hat{x}_{N-1}), \hat{\xi}_{N-1}) \\ &\quad - f(\xi_{N-2}^*, x_{N-1}^*, u_{N-1}^*(\xi_{N-2}^*, x_{N-1}^*), \xi_{N-1}^*) \\ &= f(\hat{\xi}_{N-2}, \hat{x}_{N-1}, u_{N-1}^*(\hat{\xi}_{N-2}, \hat{x}_{N-1}), \hat{\xi}_{N-1}) \\ &\quad - f(\xi_{N-2}^*, x_{N-1}^*, u_{N-1}^*(\xi_{N-2}^*, x_{N-1}^*), \xi_{N-1}^*) \\ &= F_{N-1} \cdot (\hat{x}_{N-1} - x_{N-1}^*) + G_{N-2} \cdot (\hat{\xi}_{N-2} - \xi_{N-2}^*) + G_{N-1} \cdot (\hat{\xi}_{N-1} - \xi_{N-1}^*) \\ &\quad + o(|\hat{x}_{N-1} - x_{N-1}^*|) + o(|\hat{\xi}_{N-2} - \xi_{N-2}^*|) + o(|\hat{\xi}_{N-1} - \xi_{N-1}^*|) \end{aligned}$$

故に仮定(III)を用いて

$$\begin{aligned} &E\{\hat{x}_N(0) - x_N^*(0) \mid (\hat{\xi}_{N-3}, \hat{x}_{N-2}) = (\xi_{N-3}^*, x_{N-2}^*)\} \\ &= E\{p_{N-2}^T [f(\xi_{N-3}^*, x_{N-2}^*, u_{N-2}^* + \varepsilon \zeta_{N-2}, \hat{\xi}_{N-2}) \\ &\qquad - f(\xi_{N-3}^*, x_{N-2}^*, u_{N-2}^*, \xi_{N-2}^*)] \mid (\xi_{N-3}^*, x_{N-2}^*)\} \\ &\quad + o(\varepsilon) \end{aligned}$$

故に補題により

$$\begin{aligned} &E\{H_{N-2}(p_{N-2}, \xi_{N-3}^*, x_{N-2}^*, u_{N-2}^* + \varepsilon \zeta_{N-2}, \hat{\xi}_{N-2}) \mid (\xi_{N-3}^*, x_{N-2}^*)\} \\ &\quad + o(\varepsilon) \\ &\geqq E\{H_{N-2}(p_{N-2}, \xi_{N-3}^*, x_{N-2}^*, u_{N-2}^*, \xi_{N-2}^*) \mid (\xi_{N-3}^*, x_{N-2}^*)\} \end{aligned}$$

故に $\gamma = N-2$ について (ii) が成立する。他の場合も同様にして証明出来る。 (定理証明終り)

§4. Remarks.

注意 1 §3 で証明を与えた定理は,正しい意味の最大原理ではなく, ε-maximum principle とも云うべき近似的なものである。$\{\xi_j\}$ の間に dependency がなく, $\{\xi_j\}$ が互に独立に, 而も同一分布に従うことを仮定すれば, system (2.1) の f の x および ξ に関する線型性と, optimal control の線型性 (a.s. の意味で) の条件が有れば, §3 の定理の代りに, 正しい意味での最大原理の成立することが証明出来る。所が, 我々の system では $\{\xi_j\}$ に dependency が有るために (たとえマルコフであるにせよ), 上記の仮定からだけでは最大原理は成立しない。最大原理を成立させるためには, 更に強い条件, たとえば ξ に関する線型性を仮定すれば十分である。

注意 2 §1 で, 我々の formulation が有限 stage の Blackwell model を含むことを述べたが, そのためには, 特に $n=1$ とし, $f(\xi_{j-1}, x_j, v_j, \xi_j) = x_j + r(\xi_{j-1}, v_j, \xi_j)$ とおけば, system (2.1) で $E\{x_N(0)\}$ を最大ならしめる問題は直ちに, Blackwell の有限 stage dynamic programming になる。